Amendments to the Specification

Please insert the paper copy of the Sequence Listing filed herewith following the Oath/Declaration.

Please replace the paragraph beginning at page 1, line 7, with the following amended paragraph:

This application is a continuation of U.S. Patent Application Serial Number 09/586,106, filed June 2, 2000, now U.S. Patent No. 6,720,479, which is a continuation in part of U.S. Patent Application Serial Number 09/322,478, filed May 28, 1999, now U.S. Patent No. 6,331,662, which ~~application claimed~~ claims priority to U.S. Provisional Patent Application Serial Number 60/087,125, filed May 29, 1998.

Please replace the paragraph beginning at page 63, line 21, with the following amended paragraph:

Library screening, probe preparation and PCR: Tat1 clones were obtained by screening a Landsberg erecta (La-0) 1 phage library (Voytas et al. (1990) Genetics 126: 713-721), using a probe derived by PCR amplification of La-0 DNA. The primers for probe amplification were based on the three published Tat1 sequences (DVO158, 5'-GGGATCCGCAATTAGAATCT-3' (SEQ ID NO:170); DVO159, 5'-CGAATTCGGTCCACTTCGGA-3' (SEQ ID NO:171)). Peleman et al. (1991) Proc. Natl. Acad. Sci. USA 88: 3618-3622. Subsequent probes were restriction fragments of cloned Tat1 elements, and all probes were radiolabeled by random priming (Promega). Long PCR was performed using the Expand Long Template PCR System (Boehringer Mannheim) with LTR-specific primers (DVO354, 5'-CCACAAGATTCTAATTGCGGATTC-3' (SEQ ID NO:172); DVO355, 5'-CCGAAATGGACCGAACCCGACATC-3' (SEQ ID NO:173)). The protocol used was for PCR amplification of DNA up to 15 kb in length. The following PCR primers were used to confirm the structure of Tat1-3: DVO405 (5'-TTTCCAGGCTCTTGACGAGATTTG-3'; SEQ ID NO:174) for the 3' non-coding region, DVO385 (5'-CGACTCGAGCTCCATAGCGATG-3'; SEQ ID NO:175) for the second ORF of Tat1-3 (note that the seventh base was changed from an

A to a G to make an XhoI and a SalI restriction site) and DVO371 (5'-CGGATTGGGCCGAAATGGACCGAA-3'; SEQ ID NO:176) for the 3' LTR.

Please replace the paragraph beginning at page 66, line 5, with the following amended paragraph:

If the Tat1 sequences in pDW42 and pDW99 defined retrotransposon insertions, a PBS would be predicted to lie adjacent to the 5' Tat1 elements in both clones. The putative Tat1 PBS shares similarity with the PBSs of Zeon-1 and another maize retrotransposon called Cinful (see below), but it is not complementary to an initiator methionine tRNA as is the case for most plant retrotransposons. Additionally, a possible polypurine tract (PPT), the primer for second strand cDNA synthesis, was observed one base upstream of the 3' Tat1 sequence in both phage clones (5'-GAGGACTTGGGGGGCAAA-3'; SEQ ID NO:177). We concluded from the available evidence that Tat1 is a retrotransposon, and we have designated the 3960 base insertion in pDW42 as Tat1-1 and the 3879 base insertion in pDW99 as Tat1-2. It is apparent that both Tat1-1 and Tat1-2 are non-functional. Their ORFs are truncated with respect to the coding information found in transposition-competent retrotransposons, and they lack obvious pol motifs.

Please replace the paragraph beginning at page 74, line 25, with the following amended paragraph:

The Calypso retrovirus-like elements have the same overall structure and sequence homology as the previously described Athila and Cyclops elements. The elements are ~12 kb in length; they have a 5' LTR, a PBS (Primer Binding Site), a gag protein, a pol protein, a spacer, an env-like protein, another spacer region, a PPT (Polypurine Tract) and a 3' LTR. The LTRs vary from ~1.3 to ~1.5 kb in length, and characteristically begin with TG and ended with CA. The PBS is similar to that used by the Athila and Cyclops elements; it is 4 to 6 bases past the 5' LTR and matches the 3' end of a soybean aspartic acid tRNA for 18 to 19 bases with 1 mismatch. The fact that the sequences of the Calypso primer binding sites are shared with the A. thaliana and P. sativum retrovirus-like elements, indicates that this sequence is a unique marker for envelope-encoding retroelements. The gag protein extends ~850 amino acids and encodes a zinc finger domain (characterized by the amino acid motif CxxCxxxHxxxxC; SEQ ID NO:178)

and a protease domain (characterized by the amino acid motif LIDLGA; SEQ ID NO:179). These domains are located at approximately the same positions within gag as in other retroelements. The ~600 amino acid reverse transcriptase region follows gag and has the conserved plant retrovirus-like motifs which approximate the following amino acids: KTAF (SEQ ID NO:180), MP/SFGLCNA (SEQ ID NO:181), V/I/MEVFMDDFS/WV/I (SEQ ID NO:182), FELMCDASDYAI/VGAVLGQR (SEQ ID NO:183), and YATT/IEKEL/MLAIVF/YAL/FEKFR/KSYLI/VGSR/KV (SEQ ID NO:184), respectively. The ~450 amino acid integrase domain has the plant retrovirus-like integrase motifs that approximate HCHxSxxGGH30xCDxCQR (SEQ ID NO:185) for the Zn finger as well as two other motifs that approximate WGIDFI/V/MGP (SEQ ID NO:186), and PYHPQTxGQA/VE (SEQ ID NO:187). After integrase, there is a ~0.7kb spacer then a ~450 amino acid env-like protein coding region. The env-like protein of the Calypso elements is well conserved through most of the ORF but conservation decreases toward the C-terminus. The conservation includes 2 or 3 presumed transmembrane domains and a putative RNA splice site acceptor. The env-like protein is followed by a ~2 kb spacer then a polypurine tract with the approximate sequence ATTTGGGGG/AANNT (SEQ ID NO:188). The 3' LTR starts immediately after the final T of the PPT.

Please replace the paragraph beginning at page 77, line 9, with the following amended paragraph:

Among the Calypso elements, seven have been characterized that encode env-like ORFs. These env-like ORFs form four families that have a high degree of overall sequence similarity beginning at the first methionine and continuing for three quarters of the ORF; sequence similarity falls off dramatically near the C-terminus. The amino acid sequence at the first methionine has the consensus sequence QMASR/KKRR/KA (SEQ ID NO:189), which appears to be a nuclear targeting signal, however, the program PSORT only predicts a 0.300 confidence level for this targeting role (Nakai and Horton (1999) Trends Biochem. Sci. 24:34-36). A similar sequence (ASKKRK; SEQ ID NO:190) is found at the same position in the env-like ORF of Cyclops2, suggesting that it serves a similar purpose. No other potential targeting peptide stands out from the sequence that has been analyzed so far. There is a conserved region that is

predicted to be a transmembrane domain near the center of the Calypso env-like protein and a second transmembrane domain located at variable positions near the C-terminus. These may be the fusion and anchor functions of a TM peptide. It should also be noted that five of the seven ORFs are predicted to have a transmembrane domain that is just before and includes the first methionine. This N-terminal transmembrane domain may be a secretory signal of an SU peptide. The program Tmpred estimates these transmembrane domains to be significant based on a score >500 (Hofmann and Stoffel (1993) Biol. Chem. 374:166). These three transmembrane domains are found in the Cyclops2 env-like protein at similar locations but at a reduced significance score. Another feature of the Calypso env-like ORF is the conserved splice site that is predicted to be at the first methionine by the program NetGene2 v. 2.4 with a confidence level of 1.00 (Hebsgaard et al. (1996) Nucl. Acids Res. 24:3439-3452); and Brunak et al. (1991) J. Mol. Biol. 220:49-65). There are other less preferred putative splice sites in the region, but only the splice site near the methionine is optimally placed and conserved in all seven env-like ORFs.

Please replace the paragraph beginning at page 79, line 22, with the following amended paragraph:

The degenerate oligos DVO1197 (5' GTG-CGN-AAR-GAR-GTN-NTN-AAR-YT 3' (SEQ ID NO:166) for the N terminal amino acid sequence VRKEVLKL (SEQ ID NO:167)) and DVO1198 (5' AAC-YTT-NGW-RAA-RTC-YTT-DAT-RAA 3' (SEQ ID NO:168) for the C terminal amino acid sequence VKSFDKIF (SEQ ID NO:169)) were used to amplify the Xiong/Eickbush plant retrovirus reverse transcriptase domain from genomic DNA of the following plants: New sequences were obtained from Nicotiana tabacum (Tobacco), Platanus occidentalis (Sycamore), Gossypium hirsutum (Cotton), Lycopersicon esculentum (Tomato) Solanum tuberosum (Potato), Oryza satvia (Rice), Triticum aestivum (Wheat), Hordeum vulgare (~~Barly~~ Barley), Sorghum bicolor (Sorghum), Avena sativa (Oat), Secale cereale (Rye). No sequence was obtained fro Pinus coulteri (Big-cone pine), Zea mays (Corn), Zea mays subspecies[.] parviglumis (Teosinte), and a Tripsacum species. A positive control for PCR was used to obtain previously known sequences from: Arabidopsis thaliana, Pisum sativum (pea) and three varieties (Hark 89, L85 and Williams) of Glycine max (soybean).